DOCUMENT RESUME

ED 143 497 SE 022 826

AUTHOR Ott, Mary Diederich
TITLE Observations of Cognitive and Behavioral Aspects of Physics Instruction.
PUB DATE [77]
NOTE 33p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Achievement; *Classroom Observation Techniques; *College Science; Educational Research; Higher Education; *Instruction; *Measurement Instruments; *Physics; Science Education

ABSTRACT

An observational instrument for physics instruction was developed and used in 126 class periods in courses taught by the lecture method and by an audio-tutorial method. Results indicate higher examination performance correlated positively with time spent stating and rephrasing problems, drawing diagrams on blackboards, and questioning or discussion with students. Emphasis on classroom administration and the lack of a variety of teaching behaviors correlated negatively with student exam performance. (SL)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED143497

OBSERVATIONS OF COGNITIVE AND BEHAVIORAL ASPECTS OF PHYSICS INSTRUCTION

Mary Diederich Ott

College of Engineering

Cornell University

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Mary Diederich Ott

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND THE ERIC SYSTEM CONTRACTORS."

SE 022826

OBSERVATIONS OF COGNITIVE AND BEHAVIORAL ASPECTS OF PHYSICS INSTRUCTION

Abstract

An observation instrument for physics instruction was developed and used to observe recitations in a course taught in two methods--the lecture-recitation-laboratory and audio-tutorial methods. Observations were conducted of cognitive and behavioral aspects of instruction in a total of 126 recitation periods. The reliability of the observation categories for distinguishing among teachers and between instructional methods was determined using analysis of variance. Relationships were obtained between the teacher behaviors delineated by these reliable categories and student achievement, and between teacher behaviors and student attitudes toward teachers. Furthermore, recitations were found to differ in the two methods of instruction.

OBSERVATIONS OF COGNITIVE AND BEHAVIORAL ASPECTS OF PHYSICS INSTRUCTION

Evaluations of instructional methods are incomplete without direct observation of the methods being employed (Rosenshine, 1970). Therefore, as part of an extensive evaluation of two methods of college physics instruction (Ott, 1975 and 1976; Ott and Macklin, 1975), the author conducted an observation program. The purposes of the program were (1) to determine whether student achievement and student attitudes toward teachers were related to observed differences in teacher behaviors; and (2) to specify the kinds of variation between recitations in the two instructional methods.

The two methods of instruction were lecture-recitation-laboratory (standard) instruction and audio-tutorial instruction (Postlethwait, Novak and Murray, 1972). The standard method included two fifty-minute recitations, two fifty-minute lectures, and one two-hour laboratory period per week. The audio-tutorial method included one fifty-minute recitation, no lectures, and individual study and laboratory work in the audio-tutorial learning center. The course in question is a one-semester freshman level physics course offered to approximately 550 engineering and physics majors each spring semester. Ninety-seven percent of the students were freshmen, 91% were males and 90% or more were enrolled in engineering. The course is described in detail [illegible] here (Ott 1975 and 1976).

Observations were conducted of recitations in both the standard and the audio-tutorial method, and of the lectures which were part of the standard method only. This discussion focuses on the results of the recitation observations. As will be detailed later, in this study the measure of student achievement was exam grades, adjusted for mathematics aptitude and mathematics achievement. Student attitudes toward recitation teachers were determined by means of a five-item rating scale.

The Observation Instrument

We were interested in simultaneously observing both cognitive and behavioral aspects of physics recitation (and lecture), with our primary concern being the cognitive realm. However a review of the literature did not reveal any observation instrument suitable for observing cognitive behavior in a physics classroom. Therefore, the author developed such an instrument, patterning it to some degree on the Wright-Proctor System (Wright and Proctor, 1970) which was designed for use in the mathematics classroom.

The instrument has two major components:

(1) a list of "content" categories (see Figure 1) based on the Taxonomy of Educational Objectives in the Cognitive Domain (Bloom, et. al., 1956). These categories refer to the function of the statements made by the teacher or by the students.

(2) a list of "behavior" categories (see Figure 2) based on prior observations of classroom behavior. These categories refer to the non-verbal behavior of the teacher as well as to verbal interactions between the teacher and students.

An observation period was defined as one entire recitation--about 50 minutes in duration. Every thirty seconds during an observation period, the observer noted the predominant content category and behavior category being used. Thus approximately one hundred notations of content categories and one hundred notations of behavior categories were made during each observation period.

Use of the observation instrument requires some knowledge of the terms and forms of argument used in physics. Thus a background in physics was a qualification for observers. In this study, the observers each had two or more years of college physics background. They were trained to use the observation instrument by means of audio and video tapes of physics recitations and lectures.

Design of Recitation Observations

The design for recitation observations had three factors--teacher, method, and observer. There were seven teachers, two methods, and two observers. Fourteen recitation sections taught by seven teachers (two faculty members and five teaching assistants) were observed during the course of a semester. Each instructor taught one standard and one audio-tutorial recitation section. Thus seven standard and seven audio-tutorial recitation sections were observed.

At each hour at which recitations were scheduled, students were assigned to the standard or the audio-tutorial method either randomly or by preference. (See Ott, (1976) for details of the assignment procedures.) Students were then randomly assigned to a particular recitation section, and thus to a particular teacher. The average class size was twenty students.

The seven standard recitation sections, attended by students in the standard method, met for fifty minutes twice a week for fourteen weeks. The seven audio-tutorial sections met for fifty minutes once per week. Each standard section was observed twelve times during the semester, six times by each observer. Each audio-tutorial section was observed six times, three times by each observer. Thus each of the seven teachers was observed eighteen times, for a total of 126 fifty-minute observation periods. Standard recitations were represented in 84 of the observation periods, and audio-tutorial recitations in 42.

Significance Tests

For each content and behavior observation category, the differences among teachers, among sections, and between methods were tested for statistical significance using analysis of variance. The quantity analyzed in each case was the total number of notations of a given category for each of the 126 observation periods.

Generally, the analysis focused on differences among teachers. However, for one category in which teacher by method interactions were significant (at .05), the analysis focused on differences among individual recitation sections. (A significant teacher by method interaction means that individual teachers differed in the ways they taught their audio-tutorial section relative to the ways they taught their standard section.) The sources of variation and degrees of freedom for these two forms of analysis are indicated in Figures 3 and 4.

If there is a statistically significant difference among teachers on a given observation category, the intra-class correlation coefficient for teacher effects is significantly different from zero for that category.

This occurs when the F-ratio for teacher effects for that category is significant (Rowley, 1976). Similarly, if there is a statistically significant difference among sections (or between methods) on a given observation category, the intra-class correlation coefficient for section (or method) effects is significantly different from zero. This occurs when the F-ratio for section effects (or method effects) is significant.

A convenient measure of teacher differences is the mean score for teachers for each category. For any category and teacher, this is obtained by totaling the number of observations in that category and dividing by the number of observation periods per teacher, in this case, by 18. The reliability coefficient for such a mean score measures the percentage of variation in the mean score which is accounted for by variation among teachers (Medley and Mitzel, 1958). The reliability coefficient R_t for a mean score for teachers is equal to

$$R_t = \frac{MS_t - MS_w}{MS_t},$$

where MS_t is the mean square for teacher effects and MS_w is the mean square for error (Rowley, 1976).

Similarly, the mean score for a category for sections is the total number of notations in that category for each section averaged over the number of times each section was observed. The reliability coefficient R_s for a mean score for sections is equal to

$$R_s = \frac{MS_s - MS_w}{MS_s},$$

where MS_s is the mean square for section effects and MS_w is the mean square for error. R_s gives the percentage of variation in a mean score which is accounted for by variation among sections. Here R_s is the reliability coefficient based on an average of $n_o \approx 9.0$ observation periods per section.

(See Rowley (1976) for the definition of n_o, the average number of observation periods.)

Finally, the mean score for a category for methods is the total number of notations in that category for each method averaged over the number of times each method was observed. The reliability coefficient R_m for a mean score for methods is equal to

$$R_m = \frac{MS_m - MS_w}{MS_m},$$

where MS_m is the mean square for method effects and MS_w is the mean square for error. R_m gives the percentage of variation in a mean score which is accounted for by variation between methods. Here R_m is the reliability coefficient based on an average of n_o=56 observation periods per method.

Results of Reliability Determination

Using these procedures, we determined for which categories the differences among teachers or sections and between methods were statistically significant. The level of significance was set at .05. Table 1 presents the average number of notations in each content and behavior category for an observation period and the mean squares used to determine the reliability coefficients of the mean scores. Table 1 also indicates which categories had significant F-ratios and thus intra-class correlation coefficients significantly greater than zero for distinguishing among teachers, sections, or methods. Note that content categories 8, 9, and 10 (structuring, specializing, and generalizing) were combined in the analysis due to the similarity in the cognitive level of these categories and to the small number of observations in each category.

Table 2 presents the percentages of observations in each content or behavior category which was found to be reliable, and the reliability coefficients for the mean scores based on an average of n_o observation periods. Results are grouped as follows:

(1) categories which give significant teacher differences
   (a) content categories
   (b) behavior categories
(2) the (content) category which gave a significant section difference;
(3) the (content) categories which gave significant method differences.

Thus content categories 4 (stating a problem), 7 (deriving a formula physically), and 11 (other), and behavior categories 0 (none), A (writing), C (drawing a diagram), F (demonstration), G (discussion with one student), and H (discussion with a group), reliably distinguished among teachers. Content category 1 (administration) reliably distinguished among sections. Content category 0 (silence) and the composite category 8, 9, 10 (structuring, specializing, generalizing) reliably distinguished among methods.

Table 2 indicates, for example, that content category 4 (stating a problem) accounted for about 8.5% of recitation observations. Approximately 82% of the variation in the mean scores for this category was due to differences among teachers (for 18 observation periods per teacher, consisting in 9 observation periods by each of 2 observers).

Relation of Observed Differences in Teacher Behaviors to Student Achievement and Student Ratings of Teachers

As has been emphasized by Rosenshine and Furst (1973), it is important

to try to relate the results of classroom observations to student outcomes such as course achievement and attitudes toward teachers. This step is important so that the field of education can eventually accrue evidence as to the value of various teaching characteristics, both cognitive and affective. At present only a small number of affective and cognitive characteristics have consistently been found to be related to student achievement.

Teacher Behaviors and Student Achievement

When considering the relationships between teacher behaviors and student achievement, one must keep in mind the limits of the present experimental study. First, although the observation program was extensive, a relatively small group of teachers was involved in the study, seven teachers in fourteen recitation sections. This limitation resulted from the fact that all observations had to be of teachers teaching the same course. Second, recitation was only one instructional component of the course. Third, all results are correlational rather than causal. This study did not attempt to alter teacher behavior and then to determine differences in student achievement. Finally, our measure of student achievement was fairly narrow. Due to possible differences in grading practices for the recitation portion of the grade, and to the fact that lab work was fairly independent of recitation work, we used exam grades as the measure of achievement in this part of the analysis. All students took the same prelim and final exams.

Within these narrow bounds we analyzed the correlation between teacher behaviors and student achievement. As the measure of teacher behavior we used the total number of observations in each of those categories found to reliably distinguish among teachers or among sections for each recitation section. Section totals for categories are given in Table 3 for standard

sections and in Table 4 for audio-tutorial sections. In Table 4, all totals are multiplied by a factor of two since these sections were observed half as often as were the standard sections. In these tables the seven teachers are identified by Roman numerals I through VII.

As the measure of student achievement, we used exam grades, adjusted for SAT math scores and math pretest scores. For each student we formed a total exam grade from the average of the two prelim grades plus the student's final exam grade. The average value of this total exam grade for each recitation section is given in the first column of Table 5. Using analysis of covariance, we obtained adjusted values of the total exam grade for each recitation section. The covariates were the SAT math scores and the math pretest scores. The adjusted total exam grade, the average SAT math score, and the average math pretest score are reported for each section in Table 5.

Due to the fact that the same teachers taught one standard and one audio-tutorial recitation section, we analyzed the data separately for the standard sections and for the audio-tutorial sections and then compared the results of these two analyses. For each recitation section and for each observation category found to be reliable for distinguishing among teachers or sections, we obtained the correlation coefficient assessing the degree of relationship between total numbers of observations in categories and adjusted exam totals. These coefficients are listed in columns one and two of Table 6. Finally, we averaged the two correlation coefficients for each observation category using Fisher's r to z transformation. These average correlation coefficients are listed in the third column of Table 6.

The correlation coefficients listed in Table 6 indicate trends in the data which are only suggestive due to the small number of teachers involved.

However, the results are in general agreement with earlier research. Six correlation coefficients had the same sign in both groups of recitation sections and were moderately large. Content category 1 (administration) and behavior category 0 (no other behavior) were negatively correlated with adjusted exam scores. Four categories, 4 (stating a problem), C (drawing a diagram), G (discussion with one student), and H (discussion with a group), were positively correlated with adjusted exam scores.

These data suggest a negative relationship between an emphasis on course and classroom administration (content category 1) and student achievement on exams. Low usage of behaviors listed in the behavior categories was also negatively correlated with student achievement. This type of teaching (behavior category 0) would be characterized by a large amount of talking without props, writing on the blackboard, or interaction with students. These findings agree in general with the studies reviewed by Rosenshine and Furst (1971) showing positive correlations between teaching behavior variables and student achievement. An emphasis on course and classroom administration would be inversely related to "student opportunity to learn criterion material." Low usage of behaviors listed in the behavior categories would be an indication of little "variability" in teaching behaviors. On the other hand, drawing diagrams on the blackboard (behavior category C) would be an indication of some "variability" in teaching behaviors. The category of stating or rephrasing a problem (content category 4) appears to be a measure of "task-oriented and/or businesslike behavior." However, although questioning or discussion with one student (behavior category G) and with a group of students (behavior category H) were positively correlated with student achievement in this study, no consistent relationships between these types of behavior and student achievement have been found previously (Rosenshine and Furst, 1971).

Teacher Behaviors and Student Ratings of Teachers

We also related student ratings of instructors to the instructors' use of the various reliable content and behavior categories. We sought to determine whether any general pattern would emerge relating teaching behaviors or teaching styles to student ratings of teachers. (Please note that our observation program focused on certain characteristics of classroom teaching and necessarily ignored other relevant data which may affect student ratings of teachers, such as the number of times the teacher was absent from class.) The ratings used were based on those developed and reported by Hildebrand, Wilson, and Dienst (1971). Students were asked to (anonymously) rate their recitation instructors on each of the following rating scales.

The teacher:

R1. Has command of the subject, contrasts various points of view, discusses current developments, and relates topics to other areas of knowledge.

R2. Makes himself clear, states objectives, summarizes major points, presents material in an organized manner, and provides emphasis.

R3. Is sensitive to the response of the class, encourages student participation, and welcomes questions and discussion.

R4. Is available to and friendly toward students, is interested in students as individuals, and is valued for advice not directly related to the course.

R5. Enjoys teaching, is enthusiastic about his subject, and makes the course exciting.

Ratings could run from 1 to 7, with 1 indicating that the item was "not at all descriptive of the teacher," and 7 indicating that the item was "very descriptive."

Results were analyzed by section rather than by teacher. Significant differences (at .01) were found for each rating scale. Mean values of sections' ratings are given in Table 7. Sections were usually rated in about the same order on the five scales. The pattern was as follows: From his audio-tutorial section, teacher VI received the lowest or second lowest ratings of all teachers on all scales. The audio-tutorial section of teacher I and the standard Section of teacher VII rated their teachers very highly. These two sections generally gave very high ratings on each scale.

As was the case with the analysis of observation-achievement relationships, the analysis of observation-ratings relationships was limited by the small number of sections studied. Nonetheless, a type of pattern did emerge in relation to teacher ratings. The two teachers receiving the highest ratings, teacher I and teacher VII, were quite different from each other in style. Teacher I featured a lecture style (lowest on categories G and H) emphasizing problem solutions (category 7) and teacher VII featured a questioning style (category G) emphasizing problem statement (category 4). On the other hand, the teacher who received the lowest ratings, teacher VI, was in the middle on categories 4, 7, G, and H (problem statement, problem solutions, individual and group discussion). He was high on behavior category 0 which indicates that he simply talked without using any written support. Finally, he was lowest among teachers on category C which refers to the use of drawings on the blackboard.

These results appear to indicate that the teachers who used highly developed teaching styles were rated significantly higher than was the teacher who used a fairly undeveloped style of teaching.

Differences Between Methods

In addition to the differences among teachers and sections which have been discussed, reliable differences were found between recitation sections in the two methods of instruction. There were significant differences between methods in the usage of behaviors coded in content category 0 (silence) and in the combined category 8, 9, 10 (structuring, specializing, and generalizing).

Category 0 (silence) was noted more often for audio-tutorial than for standard sections. This difference is attributable to the fact that audio-tutorial sections and standard sections had the same number of quizzes although audio-tutorial sections met only half as often as standard sections. Thus a greater proportion of class time in audio-tutorial sections was devoted to quizzes, which were categorized as 0.

Category 8, 9, 10 (structuring, specializing, and generalizing) was more common in standard then in audio-tutorial recitations. Apparently teachers reacted to the fact that audio-tutorial sections had half as much recitation time as standard sections by using these higher order cognitive categories less in their audio-tutorial than in their standard sections. This result has important implications for teaching methods which decrease the amount of time in recitation. If curriculum developers are interested in maintaining an emphasis on teaching at higher order cognitive levels in such a method, they should compensate for teachers' decreased use of these higher order levels by incorporating work at these levels in other components of the instruction, such as in the materials used in audio-tutorial learning centers.

Summary and Conclusions

Observations of physics recitation instruction indicate the utility of an observation instrument which attends to cognitive and behavioral aspects of instruction. Observations indicated a broad variation in teaching styles. For example, teaching patterns varied from one in which very little teacher-student interaction occurred, to one in which about a fourth of a typical class period was devoted to teacher-student interaction.

Within the limits imposed by a limited number of sections, we have found relations between teacher behaviors and student achievement. Higher levels of examination performance correlated positively with time spent stating or rephrasing a problem, drawing a diagram on the blackboard, and questioning or discussion with one or a group of students. Emphasis on classroom administration and little usage of various teaching behaviors correlated negatively with students' exam performance.

Further, a relationship was found between teacher behaviors and student ratings. Those recitation teachers who used highly developed teaching styles received higher ratings than did the teacher who used a fairly undeveloped teaching style. Students did not have a marked preference for discussion or for a lecture style of recitation instruction. Rather, they preferred either of these types of teaching, when highly developed, to an undeveloped teaching style having neither emphasis. These results imply that recitation teachers who are interested in improving students' attitudes toward their teaching would be advised to develop a definite teaching style. They might, for example, emphasize either the exposition of material or questioning and discussion, with adequate use of written supports such as diagrams.

Finally, teacher behaviors differed in the two methods of instruction. The major difference was that recitations in the standard method included more discussion at higher order cognitive levels than did recitations in the audio-tutorial method.

Acknowledgments

David B. Macklin contributed to every phase of this research. Walter Federer and Jason Millman provided advice on experimental design and statistical analysis. Donald Dickason, Donald Burgett, and Barbara Hirshfeld provided student aptitude test scores. Catherine Clark and William Lovejoy performed invaluable service as observers of instruction. Barak Rosenshine provided valuable comments on an earlier draft of this paper. This research was supported by a grant from the Alfred P. Sloan Foundation.

References

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. Taxonomy of Educational Objectives Handbook I: Cognitive Domain. New York: David McKay, 1956.

Hildebrand, M., Wilson, R. C., and Dienst, E. R. Evaluating University Teaching. Berkeley: Berkeley Center for Research and Development in Higher Education, 1971.

Medley, D. M., and Mitzel, H. E. Application of Analysis of Variance to the Estimation of the Reliability of Observations of Teachers' Classroom Behavior. Journal of Experimental Education, 1958, 27, 23-35.

Ott, M. D. Observations of Audio-Tutorial Learning Center Activities. Educational Technology, 1975, 15, 46-49.

Ott, M. D. Evaluation of Methods of Instruction and Procedures for Assigning Students to Methods. American Journal of Physics, 1976, 44, 12-17.

Ott, M. D. and Macklin, D. B. A Trait-Treatment Interaction in a College Physics Course. Journal of Research in Science Teaching, 1975, 12, 111-117.

Postlethwait, S. N., Novak, J. D., and Murray, A. T. The Audio-Tutorial Approach to Learning, 3rd ed. Minneapolis: Burgess Publishing, 1972.

Rosenshine, B. Evaluation of Classroom Instruction. Review of Educational Research, 1970, 40, 279-300.

Rosenshine, B., and Furst, N. Research on Teacher Performance Criteria. In Research in Teacher Education. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1971.

Rosenshine, B., and Furst, N. The Use of Direct Observation to Study Teaching. In Second Handbook of Research on Teaching. Chicago: Rand McNally and Company, 1973.

Rowley, G. L. The Reliability of Observational Measures. American Educational Research Journal. 1976, 13, 51-59.

Wright, M. J., and Proctor, V. Wright-Proctor System. In Mirrors for Behavior II: An Anthology of Observation Instruments. Philadelphia: Research for Better Schools, 1970.

Figure 1: Content Categories

0 = Silence, including silence during quizzes and films

1 = Course and classroom administration (actually giving assignments, collecting homework, calling class to order, handing out papers, asking for quiet, discussing grading of exams, etc.)

2 = References to content of assignments, exams, other parts of course such as lecture or lab (to content of a prior or upcoming exam or homework assignment; to material covered in lecture; to lab work; to audio-tutorial materials; brief reference to topics of upcoming lectures or recitations; giving answers to problems without explaining them)

3 = Describing a physical system or process or a demonstration; using a diagram or a graph to illustrate a process or kind of motion

4 = Stating a problem or rephrasing a problem

5 = Defining a concept, stating a theory, presenting a model, pointing out an analogy (definitions may be presented in terms of words or equations)

6 = Applying mathematical techniques or operations without reference to physical implications or meaning (performing a mathematical operation; deriving a mathematical concept or formula; correcting a mathematical error)

7 = Deriving or explaining an equation or formula with explicit reference to physical meaning or implications; discussing units (discussions of meaning of a term in an equation or meaning of an equation; deriving a formula for a general case or a specific example, with explicit reference to the physical principles involved)

Figure 1 (continued)

8 = Organizing or structuring the material (transitional statements, e.g., saying what will be discussed and why; reviews or previews of topics)

9 = Beginning to specialize to a particular case from a more general treatment

10 = Beginning to generalize from a particular case to a more inclusive treatment

11 = Conversation only tangentially related to subject matter (Other)

Figure 2: Behavior Categories*

O = None of the others

A = Writing an equation or words on blackboard

B = Writing an equation or words on view-graph in lecture (or showing a transparency on view-graph)

C = Drawing a diagram on blackboard

D = Drawing a diagram on view-graph in lecture (or showing a transparency on view-graph)

E = Erasing blackboard, moving view-graph, or removing transparency from view-graph

F = Performing a demonstration or showing a movie

G = Questioning or discussion with one student

H = Questioning or discussion with more than one student

I = Presenting a question for use with Instant Response System in lecture

J = Discussing student responses or correcting answer for Instant Response System question in lecture

*Categories B, D, I and J apply only to lecture, not to recitation.

Figure 3. Sources of variation and degrees of freedom for categories in which teacher by method interactions were not significant.

Source of Variation	Degrees of Freedom
Teacher	6
Method	1
Observer	1
Teacher by Method	6
Teacher by Observer	6
Method by Observer	1
Teacher by Method by Observer	6
Error	98
Total	125

Figure 4. Sources of variation and degrees of freedom for the category in which teacher by method interaction was significant.

Source of Variation	Degrees of Freedom
Section	13
Observer	1
Section by Observer	13
Error	98
Total	125

Table 1. Average number of observations for an observation period and mean squares, for each category.

	Category	Average Number of Observations Per Period	MS_t	MS_s	MS_m	MS_w
	Content Categories					
*** 0	Silence	8.3			836.03	173.49
** 1	Administration	6.8		73.49		19.55
2	Assignments	4.7	15.79			23.13
3	Describing	26.2	180.35			143.74
* 4	Stating a problem	8.4	195.05			35.66
5	Defining	4.6	31.01			21.59
6	Applying math	7.3	99.20			53.98
* 7	Deriving formula physically	24.8	395.72			90.27
*** 8, 9, 10	Structuring, specializing, generalizing	8.0			331.43	32.81
* 11	Other	0.1	0.27			0.12
	Total	99.2				
	Behavior Categories					
* 0	None	37.0	542.76			188.56
* A	Writing	27.8	249.01			76.19
* C	Drawing diagram	13.0	104.61			41.66
E	Erasing	1.8	2.82			1.84
* F	Demonstration	0.7	17.21			2.57
* G	Discussion with one student	16.9	827.87			75.40
* H	Group discussion	2.0	23.70			6.46
	Total	99.2				

Footnotes to Table 1

* Intra-class correlation coefficient > 0 for distinguishing among teachers

** Intra-class correlation coefficient > 0 for distinguishing among sections

*** Intra-class correlation coefficient > 0 for distinguishing between methods

Table 2. Percentage of observations in each reliable category and reliability of mean scores based on an average of n_o observation periods.

Category		Percentage of Content or of Behavior Observations	Reliability of Mean Score	n_o
Teacher Differences			R_t	
Content Categories				
4	Stating a problem	8.5	.82	18
7	Deriving formula physically	25.0	.77	18
11	Other	0.1	.56	18
Behavior Categories				
0	None	37.3	.65	18
A	Writing	28.0	.69	18
C	Drawing diagram	13.1	.60	18
F	Demonstration	0.7	.85	18
G	Discussion with one student	17.0	.91	18
H	Group discussion	2.0	.73	18
Section Differences			R_s	
1	Administration	6.9	.73	9
Method Differences			R_m	
0	Silence	8.4	.79	56
8,9,10	Structuring, specializing generalizing	8.1	.90	56

Table 3. Observation category totals for standard recitation sections.

Category	Teacher I	II	III	IV	V	VI	VII
Content Categories							
1 Administration	51	74	98	57	57	89	75
4 Stating a problem	31	121	112	51	129	104	156
7 Deriving formula physically	407	255	221	360	255	311	282
11 Other	1	2	4	2	1	0	5
Behavior Categories							
0 None	443	463	365	475	394	500	384
A Writing	423	315	296	421	267	352	344
C Drawing diagram	189	209	143	145	182	128	142
F Demonstration	3	14	1	1	3	4	38
G Discussion with one student	67	170	244	134	268	171	282
H Group discussion	4	22	41	12	36	16	28

Table 4. Observation category totals for audio-tutorial recitation sections. (Totals have been multiplied by a factor of two for comparability with standard sections.)

	Category	Teacher I	II	III	IV	V	VI	VII
	Content Categories							
1	Administration	154	162	52	58	82	148	66
4	Stating a problem	44	104	148	74	126	100	118
7	Deriving formula physically	354	184	286	294	306	348	284
11	Other	0	0	2	2	0	0	2
	Behavior Categories							
0	None	524	598	396	426	376	604	352
A	Writing	342	234	360	288	298	372	288
C	Drawing diagram	162	166	122	170	174	92	124
F	Demonstration	6	0	6	0	2	2	26
G	Discussion with one student	72	182	252	220	316	188	346
H	Group discussion	18	22	40	8	58	18	28

Table 5. Mean achievement and trait data for each recitation section.

Teacher	Number of Students*	Total Exam Score**	Adjusted Total Exam Score**	Math SAT	Math Pretest Errors
Standard Sections					
I	17	108.06	112.59	689	3.71
II	25	129.04	129.37	672	2.48
III	27	116.22	118.31	669	2.78
IV	23	119.65	118.18	698	2.70
V	19	127.05	126.67	683	2.58
VI	18	113.53	111.58	698	2.61
VII	19	117.61	115.30	695	2.47
Audio-Tutorial Sections					
I	26	120.92	119.48	690	2.42
II	14	96.32	97.91	687	3.07
III	21	131.95	126.42	694	1.81
IV	18	107.39	114.83	661	3.67
V	19	125.08	127.74	679	3.11
VI	22	112.43	110.23	691	2.41
VII	18	111.28	110.52	690	2.67

* The criteria used for including students' data are as follows: We included data only for those students in the random assignment sections who had been originally assigned to the recitation sections in which they continued, and for those preference assignment students who actually obtained the method they preferred and did not transfer sections. Students who dropped the course or transferred sections were excluded from the analysis.

** Scores differed significantly (at .05) across the fourteen sections.

Table 6. Correlation coefficients for adjusted total exam scores vs. observation category totals

Category	Correlation Coefficients		
	Standard (n = 7)	Audio-Tutorial (n = 7)	Average*
Content Categories			
1 Administration	-.143	-.545	-.354
4 Stating a problem	.376	.224	.300
7 Deriving formula physically	-.564	.571	-.064
11 Other	.017	.176	.097
Behavior Categories			
0 None	-.170	-.619	-.422
A Writing	-.611	.548	-.046
C Drawing diagram	.663	.086	.414
F Demonstration	-.025	-.031	-.025
G Discussion with one student	.311	.202	.254
H Group discussion	.447	.604	.530

* Averages were formed using Fischer's r to z transformation.

Table 7. Mean ratings of recitation instructors.

Teacher	Maximum Number of Ratings Received*	Ratings R1	R2	R3	R4	R5
Standard Sections						
I	12	5.75	6.25	6.25	5.92	5.83
II	28	5.61	5.22	5.07	3.92	4.52
III	24	5.09	5.54	5.75	5.09	5.13
IV	19	5.00	5.95	5.47	5.00	4.74
V	12	5.08	5.75	6.33	5.17	5.42
VI	14	4.93	5.50	6.14	4.85	4.93
VII	16	6.44	6.19	6.31	6.00	5.81
Audio-Tutorial Sections						
I	22	6.27	6.73	6.36	6.18	5.96
II	13	6.08	6.08	6.00	5.50	5.69
III	19	4.50	5.47	5.79	5.07	5.00
IV	13	5.85	6.23	6.46	5.67	5.58
V	10	5.50	5.20	6.30	5.70	4.80
VI	19	4.63	4.95	4.58	4.12	4.17
VII	20	5.65	5.45	6.15	5.47	5.20

* Total numbers of responses differed for each rating. For R1, n = 238; R2, n = 239; R3, n = 241; R4, n = 223; R5, n = 237.